

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2001, 16:52:25 ; Search time 16.97 Seconds
(without alignments) 1456.222 Million cell updates/sec

Title: US-09-830-647-1

Perfect score: 3510
Sequence: 1 MNSGAMRIHSKHFGGIGIY.....SDNLTAFSSPSTSTFTG 674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwisProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187	5.1	1790	1	USO1_YEAST
2	178	5.1	3911	1	AKA9_HUMAN
3	165	4.7	1433	1	REST_CHICK
4	163.5	4.7	2230	1	GOG4_HUMAN
5	159.5	4.5	1427	1	REST_HUMAN
6	156.5	4.5	3210	1	CENF_HUMAN
7	155.5	4.4	1056	1	YNN2_YEAST
8	154	4.4	2663	1	CENE_HUMAN
9	152	4.3	1216	1	PIB1_HUMAN
10	152	4.3	1312	1	RA50_YEAST
11	152	4.3	1980	1	MY9B_NAT
12	151.5	4.3	1044	1	YAF3_SCHPO
13	151.5	4.3	2104	1	MY53_SCHPO
14	149.5	4.3	1726	1	MSPI_PLAFK
15	149.5	4.3	1726	1	MSPI_PLAFK
16	149	4.2	2245	1	MY51_DICDI
17	148.5	4.2	1172	1	CM14_SCHPO
18	148.5	4.2	1805	1	HMM2_MYGE
19	148.5	4.2	1914	1	MY9B_MOUSE
20	148	4.2	1875	1	MLP1_YEAST
21	147.5	4.2	844	1	BR01_YEAST
22	147.5	4.2	1978	1	MYHB_CHICK
23	147	4.2	1957	1	Y086_SCHPO
24	146.5	4.2	1216	1	PIB1_NAT
25	146.5	4.2	1257	1	RBB1_HUMAN
26	145.5	4.1	756	1	Y328_MYGE
27	145.5	4.1	1630	1	MSPI_PLAFK
28	145.5	4.1	1639	1	MSPI_PLAFK
29	144.5	4.1	1031	1	KINH_STRPU
30	144.5	4.1	1225	1	SMC1_YEAST
31	144.5	4.1	1226	1	YCS3_YEAST
32	144.5	4.1	1290	1	XCPC_XENLA
33	143.5	4.1	1616	1	P200_MYGE

34	143	4.1	1087	1	AKA9_RABIT	Q28628 oryctolagus
35	142.5	4.1	935	1	KINH_SYRNA	Q43093 syncephalus
36	142.5	4.1	1216	1	PIB1_BOVIN	P10894 bos taurus
37	142.5	4.1	1453	1	MYH3_HUMAN	Q9ubc1 homo sapien
38	142.5	4.1	1513	1	STU1_YEAST	P38198 saccharomyc
39	142.5	4.1	1679	1	Y109_YEAST	P40457 saccharomyc
40	142	4.0	1701	1	MSPI_PLAFK	P13819 plasmodium
41	142	4.0	1701	1	MSPI_PLAFK	P08569 plasmodium
42	142	4.0	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
43	142	4.0	1937	1	MYHB_HUMAN	P13535 homo sapien
44	141.5	4.0	1972	1	MYHB_RABIT	P35748 oryctolagus
45	141	4.0	853	1	YCG1_YEAST	P25588 saccharomyc

ALIGNMENTS

```

RESULT 1
USO1_YEAST STANDARD: PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USO1.
CN USO1 OR INT1 OR YD1058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A;
RX MEDLINE=91183402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, uso1, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RC Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54378; CAA38253.1; -
DR EMBL: L03188; AAB00143.1; -
DR EMBL: U53668; AAB6659.1; -
DR PIR: A38455; A38455.
DR HSSP: P80220; IDIP.
DR SGD: S0002216; USO1.
DR InterPro: IPR002017; Spectrin.

```

Query Match 5.3%; Score 187; DB 1; Length 1790;
 Best Local Similarity 19.9%; Pred. No. 0.016;
 Matches 150; Conservative 133; Mismatches 273; Indels 198; Gaps 31;

Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 KW DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B2169FD4818 CRC64;

2 NSGAMRIHSGHFGGIGOVNENKRNPSLKTNDREKSKCPKMGKVFYDLPSTVIS 61
 978 NESLIKAVESKNESSIQSLNKNKIDSMOSEN-----FOIERGSIERN 1023
 62 -EKLONKIDGGRVEEFLSK-----DISYLISNKKKAPFA--QTIGRISPVSP 108
 1024 IEQKTIISDLEQTKREIISKDSKDEYEQISLMEKLETATNDENVKISIELTKT 1083
 109 ESAYTAEFTSPHSHSGSPSPDVCLSRGLKVEAIDHDFISNSILSNALSMGVK 168
 1084 REELAEALAA-----YKMLKELETKLETS-----EALKE--VENE----- 1119
 169 ILHDDIRYIEOK-----KKELYLKKSSTVROGGRVSGMGKRTGRLKPPVKED 224
 1120 -EHKEREKIOLEKAEETEQOOLNSLANBSEKHEHDLAOLKTEPOIANKEQYNEE 1178
 225 MSQLYRFYQILNMFPIYSTQKCPSPDYD---KSSMGKQTQVKRLIQDGRKYGT 281
 1179 ISQND---ELTSTQOENESIKRKNDELBEVAMKSTSEOSNKL-RSEID----- 1226
 282 SIQQLKERRKKCYC-----ECCLQKYE--DLETHLLSDQH 315
 1227 ALNLQIKELKKKNETNEASLSIKSVSEETVKIKELQDCEKNEKEVESELDKASED 1286
 316 RNF-----AQSNQYVVDIVSKLVEFVEYEDTPKKRKRIYSVGSLSVSAVLKTEQ 371
 1287 KNSKYLLEQKESKEKEEEDAKTELKIQLEKTNLKAKEKSESLSR-----LKTSS 1341
 372 KEKVELQIHISQKDCQDDDTTVEQNFLYKETOETKKLL-----FTSEPIPHPSNELG 426
 1342 EER-----KNAEQQLKELKNEIQIKNOAFKEREKRLNLEGGSTITQOYSKINTLDE 1393
 427 NEKKSNCMSLSTAEEDITRONFOLPLH-----KNKQECILDSERTLS----- 470
 1394 LIRQNNNEKKAIEDNTSELEKSLSNDELBEKNTKTSIQDITLYKRIKTRNDEK 1453
 471 -----ENDLELA-----VDHKCNIO-----ASV 490
 1454 LLSIERNNKDLLESLLEQLRAOESKAVEEGKLKLEESSKKAELKSKEMAKLEST 1513
 491 HVSDFTSDNGSQPKOKSPTVFLPAKDIKEKDLHSIFTHDSGLIT--INSQD-----H 542
 1514 IESNETELKSMETIKRSDEKLEQSKSAEEDINNNIHNESINSESDIELELKS 1573
 543 LTVQAK--APFHTPPEPEPCDFKKNMDSLPGKIHKKVLIILGRNKENLEFPNAFDK 600
 1574 LRIEAKGSGLEFVVKOLNNAOEKI-----RINAENVTL--KSLIEDIERELK--DKOA 1624
 601 EF--ITQENRICSSPVQSL--LDLFQTSSEKSE 630
 1625 EIKSNOBEKELLTSRLKLEQLDSTOOKAKRSE 1658

RESULT 2
 ID AKAP9_HUMAN STANDARD; PRT; 3911 AA.
 AC O99996; O99004; O99083; O99612; O14869; O43355; O94895; O99688;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE A KINASE ANCHOR PROTEIN 9 (PROTEIN KINASE A ANCHORING PROTEIN 9)
 DE (PKAP9) (A-KINASE ANCHOR PROTEIN 450 KDA) (AKAP 450) (A-KINASE ANCHOR
 DE PROTEIN 350 KDA) (AKAP 350) (HGAKAP 350) (AKAP 120 LIKE PROTEIN)
 DE (HYPERTON PROTEIN) (TOTITAO PROTEIN) (CENTROSOME- AND GOLGI-LOCALIZED
 DE PKA-ASSOCIATED PROTEIN) (CG-NAP).
 GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RX MEDLINE=98151389; PubMed=9482789;
 RA Lio J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
 RA "Vollao, a novel protein of neuromuscular junction and brain that
 RA interacts with specific splice variants of NMDA receptor subunit
 RA NR1.";
 RA J. Neurosci. 18:2017-2027(1998).
 RL [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
 RC MEDLINE=99219864; PubMed=10202149;
 RA Mltcak O., Skralnegg E.S., Kreyer G., Bornens M., Tasken K.,
 RA Jahnson T., Oestervik S.,
 RA Cloning and characterization of a cDNA encoding an A-kinase anchoring
 RA protein located in the centrosome, AKAP450.";
 RA EMBO J. 18:1838-1866(1999).
 RL [3]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=99287934; PubMed=10358086;
 RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
 RA "Characterization of a novel giant scaffolding protein, CG-NAP, that
 RA anchors multiple signaling enzymes to centrosome and the golgi
 RA apparatus.";
 RA J. Biol. Chem. 274:17267-17274(1999).
 RL [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC Kemmer M.A., Dells S., Schwarz U.;
 RA "Cloning of Hyperion.";
 RA Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
 RC TISSUE=Gastric parietal cell;
 RX MEDLINE=99115654; PubMed=9915845;
 RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
 RA Trotter K.W., Mligam S.L., Goldentring J.R.,
 RA "AKAP350, a multiply spliced, protein kinase A-anchoring protein
 RA associated with centrosomes.";
 RA J. Biol. Chem. 274:3055-3066(1999).
 RL [6]
 RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
 RC TISSUE=Lymphoblast;
 RA Hinds K., Suterer C., Becker M., Hawkins M.;
 RA Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.
 RL [7]
 RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
 RC TISSUE=Lung;
 RA Mligam S.L., Goldentring J.R., Schmidt P.H.;
 RA "AKAP350: A multiply spliced family of proteins with centrosomal
 RA association.";
 RA Submitted (Sep-1998) to the EMBL/GenBank/DBJ databases.
 RL [8]
 RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).

CC TISSUE-Brain;
 RX MEDLINE-99087487; PubMed-9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:277-286(1998).
 RN [9]
 RP SEQUENCE OF 17-1800 FROM N.A.
 RA Wu X., Graves T., Bradshaw H.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
 CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
 CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
 CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
 CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
 CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
 CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
 CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
 CC -1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
 CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
 CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PRC EPSILON.
 CC CYTOPLASMIC IN PARIETAL CELLS.
 CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-
 CC NMP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
 CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
 CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
 CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
 CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
 CC -1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
 CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AJ131693; CAB40713.1; -
 DR EMBL: AB019691; BAA78718.1; -
 DR EMBL: AJ010770; CAA09361.1; -
 DR EMBL: AF026245; AAB86384.1; -
 DR EMBL: AF083037; AAD22767.1; -
 DR EMBL: AC004013; AAB96867.1; ALT_FRAME.
 DR EMBL: AF091711; AAD39719.1; -
 DR EMBL: AB018346; BAA34523.1; -
 DR EMBL: AC000066; AAC60380.1; ALT_FRAME.
 DR MIM: 604001; -
 KM Colled coil; Alternative splicing; Polymorphism.
 FT DOMAIN 2567 PRA-RII SUBUNIT BINDING DOMAIN.
 FT 164 914
 FT DOMAIN 1022 COILED COIL (POTENTIAL).
 FT 1100 1185 COILED COIL (POTENTIAL).
 FT DOMAIN 1253 COILED COIL (POTENTIAL).
 FT 1336 1392 COILED COIL (POTENTIAL).
 FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
 FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
 FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
 FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
 FT 2603 2776 COILED COIL (POTENTIAL).
 FT 3065 3092 COILED COIL (POTENTIAL).
 FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
 FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).
 FT DOMAIN 3726 3730 POLY-LEU.
 FT 203 292 GLN-RICH.

FT DOMAIN 321 1010 GLU-RICH.
 FT 1846 2772 GLU-RICH.
 FT VARSPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT VARSPLIC 1637 1642 QLOEET -> LATRED (IN ISOFORM 4).
 FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 4).
 FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 3).
 FT VARSPLIC 2175 2183 SADFOKE -> Q (IN ISOFORM 6).
 FT VARSPLIC 2895 2907 VGFENKCFSTLC -> GSSITELASDAVOTREICSS
 FT (IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).
 FT VARSPLIC 2895 2948 MISSING (IN ISOFORM 5).
 FT VARSPLIC 3901 3911 STOFHAGMR -> ALSITTSWOHSARPTAPLEFETLSH
 FT SLG (IN ISOFORM 6).
 FT VARIANT 1347 1347 K -> KO.
 FT
 FT CONFLICT 76 76 E -> Q (IN REF. 3).
 FT CONFLICT 475 475 M -> I (IN REF. 3).
 FT CONFLICT 554 554 E -> G (IN REF. 3).
 FT CONFLICT 638 638 R -> S (IN REF. 3).
 FT CONFLICT 663 663 N -> S (IN REF. 3).
 FT CONFLICT 913 913 H -> N (IN REF. 3).
 FT CONFLICT 956 956 K -> N (IN REF. 3).
 FT CONFLICT 980 982 OKH -> PKP (IN REF. 1 AND 2).
 FT CONFLICT 997 997 Q -> P (IN REF. 1 AND 2).
 FT CONFLICT 1001 1001 Q -> P (IN REF. 1 AND 2).
 FT CONFLICT 1020 1020 N -> D (IN REF. 3).
 FT CONFLICT 1028 1028 V -> E (IN REF. 3).
 FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).
 FT CONFLICT 1703 1703 N -> T (IN REF. 3).
 FT CONFLICT 1707 1707 V -> G (IN REF. 3).
 FT CONFLICT 1802 1803 MISSING (IN REF. 5).
 FT CONFLICT 1843 1843 A -> P (IN REF. 3).
 FT CONFLICT 1956 1956 I -> V (IN REF. 3).
 FT CONFLICT 2027 2027 V -> D (IN REF. 5).
 FT CONFLICT 2157 2158 EI -> HE (IN REF. 7).
 FT CONFLICT 2169 2169 E -> V (IN REF. 3).
 FT CONFLICT 2514 2514 L -> R (IN REF. 3).
 FT CONFLICT 2851 2851 I -> N (IN REF. 8).
 FT CONFLICT 2957 2957 E -> D (IN REF. 3).
 FT CONFLICT 2983 2983 P -> S (IN REF. 3, 7 AND 8).
 FT CONFLICT 3087 3087 Q -> H (IN REF. 3).
 FT CONFLICT 3218 3218 Q -> H (IN REF. 3).
 FT CONFLICT 3307 3309 ESE -> OSO (IN REF. 3).
 FT CONFLICT 3751 3751 P -> A (IN REF. 3).
 FT CONFLICT 3833 3833 T -> S (IN REF. 3).
 FT SEQUENCE 3911 AA; 453664 MM; 3FB1CB1819B4/AA CRC64;
 SQ

Query Match 5.1%; Score 178; DB 1; Length 3911;
 Best Local Similarity 21.1%; Pred. No. 0.13;
 Matches 115; Conservative 104; Mismatches 184; Indels 142; Gaps 27;
 QY 179 IEQKKRELYLLKKSSTVARDGKRGVSGAOKTRTGRLKPKFVVEDMSQLYRPFYLQITN 238
 DB 351 IEEKKTELEKDKLTADKLGLGEOIVQKNO--EIKNMKLELINSKOKESSEIEIKO 408
 QY 239 MEFIVYSIQK---PCSPEDVNDKPSMOKOTQVK---LRQIDGDKYGGSTISDL 285
 DB 409 IMGVEEELQKRNHKSOFETDVGQREBQETORLEQRLD-EMVGOIIVMKEILIRQ 467
 QY 286 ---OLKERR--KKGYCECLOKEDLETH-----LSPQHNRFQSNQ----- 323
 DB 468 HMAQNEKMKTRKKGEMENALNSYSNITVNEODIKLMNAVALNLIKLDITNSQKEREKEE 527
 QY 324 -----YQVVDIVSKLVF--DFVEYBKDPYPKKRIKY-----SVGSI,SPVSASV 365
 DB 528 IGLILIEKCALQROLEDLYEELSFBRQIQRARQTIABGESSLNEAHKSLSTVEEDLKAEI 587
 QY 366 IAKTQKEKVELQHSIQKDCQEDDTT-----VKQONFL---YKETOEPF-----KK 408
 DB 588 VSASESRKELELKH-----EAEVTNVIKLEMLEKEKNVAILDMABESQAELERLRLQ 640
 QY 409 LFIIE-----PIPHPSN-----ELRGINEMKSNKCSMLSTVAEDDI 444

DB 641 LILFSEBELSKLEDEIEHRIINIKLNDICGHIYKQOIGLONEMSOKEITMOFEKNDL 700
 445 RQNTOLPLH-----KROKCIID--ISEHTISENDE-ELRV-----DHYKCNIOASVHS 493
 701 ITRKONOLLEISKIKDLOOSLVNSKSEMTLOJNELOKEIILROEKEKGTLEOQVEL 760
 494 DESPDNGSGPKOKSPDTVLEPAKDKERDLSIFTHDSGLITINS--SOEHLTVQAKAPF 551
 761 QLTLELLEKQMKERK-----DLQEK-----PAQLEAENSILKDEKKTLEDMLEKI 805
 552 HTPEPEENEDDFNMOSLPS-----GKIHRYKVIILGRN--RKENLEPNAPFDR-RTEPI 603
 806 HTPVSOEERLIF--LDSIKSKSDSWERKEIEIIEFNEDLKQOCIOLEIEKORNTFS 863
 604 TOEEN 608
 864 PAENK 868

RESULT 3
 REST_CHICK STANDARD: PRT: 1433 AA.
 ID 042184: 042228; 057563; 057564;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RESLIN (CYTOPLASMIC LINKER PROTEIN-170) (CLIP-170).
 GN RSN.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98137792; PubMed=9469933;
 RA Gripatic L., Volosky J.M., Keller T.C. III;
 RT "Cloning and expression of chicken CLIP-170 and reslin isoforms.";
 RL Gene 206:195-206(1998).
 [2]
 RP SEQUENCE OF 17-1139 FROM N.A. (CLIP-170(11) AND CLIP-170(11+35)).
 RC TISSUE=pectoralis muscle; III;
 RA Gripatic L., Keller T.C. III;
 RT Identification and expression of two novel CLIP-170/Reslin isoforms
 expressed predominantly in muscle.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 "TAAV LINKS ENDOCYTIC VESICLES TO MICROBULLES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CYTOSKELETON (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
 ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to jlicse@sib-stb.ch).
 CC -----
 CC EMBL: AF014012; AAC60344.1; -
 DR EMBL: AF020764; AAC60345.1; -
 DR EMBL: AF045650; AAC03547.1; -
 DR EMBL: AF045651; AAC03548.1; -
 DR InterPro: IPR000938; CAP-GLY.
 DR InterPro: IPR001878; zn1_CCHC.
 DR Pfam: PF01302; CAP-GLY; 2.
 DR SMART: SM00343; zn1_C2HC; 1.
 DR PROSITE: PS00845; CAP-GLY_1; 2.
 DR Cytoskeleton: Microtubules; Colled coil; Alternative splicing.
 FT DOMAIN 79 121 CAP-GLY 1.

FT DOMAIN 144 207 SER-RICH.
 FT DOMAIN 235 277 CAP-GLY 2.
 FT DOMAIN 305 332 SER-RICH.
 FT DOMAIN 351 1353 COILED COIL (POTENTIAL).
 FT DOMAIN 1414 1427 CCHC-BOX.
 FT VASPLIC 458 492 MISSING (IN SHORT ISOFORM).
 FT VASPLIC 458 492 TOTLEHARKELOSLFEKTKADKLORELEDR ->
 FT VASPLIC 803 803 RKRISDEPN (IN ISOFORM CLIP-170(11)).
 FT VASPLIC 803 803 T -> RKRISDEPN (IN ISOFORM CLIP-170(11+35)).
 FT VASPLIC 458 458 T -> RKRISDEPN (IN ISOFORM CLIP-170(11+35)).
 FT CONFLICT 309 309 K -> R (IN REF. 2; AAC03547).
 FT CONFLICT 440 440 E -> V (IN REF. 2; AAC03548).
 SO SEQUENCE 1433 AA; 161026 MW; 5631CE8683498E23 CRC64;

Query Match 4.7%; Score 165; DB:1; Length 1433:
 Best Local Similarity 18.7%; Pred No. 0.17; Mismatches 303; Indels 144; Gaps 32;
 Matches 135; Conservative 139;

DB 11 KGFHGGIOYKN-----EKNRPSIKSLKTDNRPKSKCPKNGKVPYL-----DLPS 57
 771 KANSEKLEIOLKLESOLOAKERQIONLETEKVSN-----LTKELQKEKGLDLEKINLSA 825
 58 VT-ISEKLOKDKI-----DLCGRVEEFLSKDISYILSNKKEAKPAOTLGRISFVP 106
 826 VNOVKOSLEKEILOLKEKETSAVDGAENAGRAMQETINKL--NOREQOPALMSELEOLK 883
 107 SPESAYTAETTSPPSHDGSSEKSPDTVCLSRGLYERAKINDHDFIPNSILSNALSWG 166
 884 S--NLTVMET-----KLKERE-----EREQOLTEAKVKLENDIA--ELIMSSGDS 925
 167 VKLIHIDDIYYIEOKKELYL-LKSSSTSVROGKRVSGAQKRTGKPKPKPYEDM 225
 926 AQMKMNDLELRLEROLEIOLELTKANERAVOLAKKVEQTAQKEOSQOETLTHQDEL 985
 226 SOLRPVYLOLTMPPIVINSIQPCSPDVDPKPSMOKOTQVRLRIQT-----DGGKY 278
 986 KKMDD--QLTDM-----KKMETSQNYKDLQAKYERETSEMITKIDADIK 1029
 279 GGTSTIOLKKEKKKKKGCYCCLOKYEDELTHLSEQRNFAQSNQVQVDDIVSLYVDF 338
 1030 GFKNLDAEALK-----AAQKMDLELTQ-ABELLKAQAEQAKDKRAEVELQTM--EK 1081
 339 VEIKDTPKKRKRIKIVSVSLSPVSAVLAKTEQKEKVELQHSIQDCQDDPTVYEQNFL 398
 1082 VTKEKDAIHOEKTE--TLASLENS--KOTNEKLQNELMLQNKLNKNEBELIKSKELL 1135
 399 --YKETOETEKKLFISEPIHPSPNEIENKMSKCSMLSTAEADDIRQNTOLPLHK 455
 1136 NLENKKVEELKKEFEELKLAQAQSQALALQDEENVKLABELGRSDEVTs-----HO 1188
 456 NKOCILDISENTLSENDEELR-----VDHYKCNIOASVHSDFS--TDNSGQ 503
 1189 KLEP-----ERSVLLNNOLLEKKKRESLTKKEIDEERASLOKS--ISDSALITQKDEL 1240
 504 PKOKSDVLEPPARDLKEKDIHSIF-THDSGLITINSQEHLYVOAKAPRPPPEPNCD 562
 1241 EKLNETTVARGENASAKTIGSVVTLSEDKLEEKVKNLEQKLEKAEQDFLVTS-- 1297
 563 FKNNDSPSGKTIIRKVKIILGRNRENLEPNAEF-----DKRTEFTQENRNICSSP 614
 1298 -----PSGDIAN--LLODESADKQOEIDPLNSVIVDQORNEELNLKIQMCEAA 1347
 615 VOSLLDFORSEKSEKSEFLGFTSYTEK--SGICNVLDIWEENSNDNLTAFFESPSTSTFT 672
 1348 LNGEE--ETINVOSEEBGLSKTTRLPDLCIGCFDLDHEDCPT-QAQMLEPPHSHYH 1404
 673 G 673
 DB 1405 G 1405

GN RSN.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Peripheral blood monocytes;
 RA MEDLINE=92289675; PubMed=1609342;
 RA Bille G., Delaite J., Bruessen J., Richener H., Asselbergs F.A.M.,
 RA Celletti N., Sorg C., Orlan K., Tarsany L., Wiesendanger W.,
 RA de Wolf-Petersen C., Shipman R.,
 RA Reslin: a novel intermediate filament-associated protein highly
 RA expressed in the Reed-Sternberg cells of Hodgkin's disease.;
 RT EMO J. 11:2103-2113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92405160; PubMed=1356075;
 RA Piarre P., Scheel J., Rickard J.E., Kreis T.E.;
 RA "CIP-170 links endocytic vesicles to microtubules.";
 RT Cell 70:887-900(1992).
 CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CC CYOSKELETON.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
 CC OF HODGKIN'S DISEASE.
 CC -1- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
 CC
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 CC
 CC EMBL: X64838; CAA6050.1; .
 CC EMBL: M97501; AAA35693.1; .
 CC PIR: S22695; S22695.
 CC MIM: 179838; .
 CC InterPro: IPR000938; CAP-GLY.
 CC InterPro: IPR001878; znf_CCHC.
 CC Pfam: PF01302; CAP-GLY; 2.
 CC SMART: SM00343; znf_C2HC; 1; 2.
 CC PROSITE: PS00845; CAP-GLY_1; 2.
 CC Cytokeleton: Microtubules; Coiled coil; Alternative splicing.
 CC KW DOMAIN 60 125 CAP-GLY 1.
 CC FT DOMAIN 143 204 SER-RICH.
 CC FT DOMAIN 214 279 CAP-GLY 2.
 CC FT DOMAIN 304 331 SER-RICH.
 CC FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1408 1421 CCHC-BOX.
 CC FT VASPLIC 457 491 MISSING (IN SHORT ISOFORM).
 CC FT COMPLY 1069 1069 P->E (IN REP. 2).
 CC SEQUENCE 1427 AA; 160989 MW; 0AA6166DB94254EB CRC64;

Query Match 4.5%; Score 159.5; DB 1; Length 1427;
 Best Local Similarity 17.3%; Preg. No. 0.34;
 Matches 127; Conservative 137; Mismatches 295; Indels 175; Gaps 28;
 QY 11 KGRHGGIOGVN-----EKNRPSLSKLTNDNRPKSKCKPLMGKVEYLLDLPVYTISEKQLQ 65
 DB 770 KASSEGSKEMKKRIKROOLEAEMKQIKHLETKNAESKASSTRLOGRLEKULNLOENLS 829
 QY 66 KDIDLGAREEFLSKDISYLSNKKKFAFQTLGRISPVSPSPASATVA-----ETTSF- 119
 DB 830 E-----VSQVKETLEKLEQL-----KEKFAE-----ASFEAVSVQORSQGFVKVL 870
 QY 120 HPSHDGSSFPSTVCLSGKLLVKAIKKDHDFIPNSILSNLSMVGKILHIDIRYI 179

DB 871 HOKEDFNMSSDLERENLADMEAKFREK----- 902
 QY 180 EOKKKELYLLKSSSTSVRDGGRVNG-SGAOKRTPTG---RUKRPFVKVEDMSQLYRPPYLO 235
 DB 903 EREBOLIKAKEKLENDIAEIMKMGDSSQLTKMDELRLKE-----RDVEL-----OLK 953
 QY 236 LT-----NMPFINSIQKPCSPEDVDKPSMOKOTQVRLRIQDGDGKYGSTISQLOLKEK 291
 DB 954 LTKANENASPLQKSIEDMTVKAESQOEAAKKHKEEKELER-----KLSDLEKKMET 1006
 QY 292 KKGCECCQCKYEDLETHLSBOHRNFAOSNOYOYVD-----DIVSLVDPFVEYE 342
 DB 1007 SHNOCQELAKRYE-RATSETKTKHBEILQNLQKTLDPTEDEKLGARENSGLQLELELR 1065
 QY 343 KDTPKKRIKISVGSLSFVSASVYLKTTQKKEVLELOHISQKOCQEDDTTVEQNFVKET 402
 DB 1066 KQADKAAQOTADDAQOTMEQMTKEKTEETTLASLEDTKQTKAKLQNELDTLKNLKNVEE 1125
 QY 403 QETEKLLFT-SEPIPHSPNELRLGINKSNKCSSTVAE-----DDIKONFTQLPLH 454
 DB 1126 LNRKSKELLVYENQKMEEFKELETLKQAQAKSQOULSAQENVKLAELGSRQEVYSH 1185
 QY 455 KNOECILDISHTLSENDLELR-----VDHYKNIOASVHV-SDFSTDNGSQOP 504
 DB 1186 QKLEE-----ERSVNLNQLLEMKKRESKFIKQADEKASLQKSISSITSAITLKEDAELE 1239
 QY 505 KQKSDTVLPFAKDLKEKDLHSIF-THDQGLITINSQOHLTVQAKAPFHTPEEPNECDF 563
 DB 1240 KLRREVTVLNGENASAKSLHSVQTLSDQVKLELKVKNLLEQLK-----EN 1286
 QY 564 KNDLSLSCGIRHKVYKILGRNRKENLEPNAEFKRT-----EFIT-----QENNRICS 612
 DB 1287 KROLSSSG-----NDTQADEDERAESQIDFLNSVYVDLQKRNODLK 1330
 QY 613 SPVQSL-----LDLPQTSSEKSEFLGFSTYKSGICGVNVDIYEENSNNLLT- 660
 DB 1331 MKYVMEALMANGDNDLNNYSDQGEK-----SKKKPLRFDICDDGDLHDREDCQTQ 1385
 QY 661 -APFSSPSTPTFG 673
 DB 1386 AQMSDEPPSTHNG 1399

RESULT 6
 CENF_HUMAN STANDARD; PRT; 3210 AA.
 AC P49454; Q13246; Q13171;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CENP-F KINECHORE PROTEIN (CENTROMERE PROTEIN F) (MITOSIN) (AH
 DE ANTIGEN).
 GN CENP-F.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RA MEDLINE=95348175; PubMed=7542657;
 RA Liao H., Winkler R.J., Mack G., Rattner J.B., Yen T.J.;
 RA "CENP-F is a protein of the nuclear matrix that assembles onto
 RA kinetochores at late G2 and is rapidly degraded after mitosis";
 RT J. Cell Biol. 130:507-518(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
 RA Jones D., Yang-Feng T.L., Lee W.-H.;
 RA "Characterization of a novel 350-kilodalton nuclear phosphoprotein
 RA that is specifically involved in mitotic-phase progression";

CC -1- SIMILARITY: TO S.POMBE SPAC20G8.09C AND AN A.AMBISEXUALIS
 CC HYPOTHEICAL PROTEIN (AC P54008).
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 246843; CAA6893.1; -
 CC DR EMBL: 271408; CAA96014.1; -
 CC DR SGD: S0005076; YNL132W.
 CC Hypothetical protein; ATP-binding.
 CC NP_BIND 283 290
 CC FT SEQUENCE 1056 AA; 119347 MW; 76721ED0867ED618 CRC64;

Query Match 4.4%; Score 155.5; DB 1; Length 1056;
 Best Local Similarity 21.0%; Pred. No. 0.37; Indels 235; Gaps 37;
 Matches 151; Conservative 98; Mismatches 235; Indels 235; Gaps 37;
 CC
 CC 19 QVNEKNEPRLSKLT-----DNREPKSKCPLMGKVFYDLPVWISF-KLOKDIKDL- 71
 CC 429 QLRNONNTSGHSTQTAIVVSRDKERDS-----HLHSQSRQLRETSIDELRVAP 478
 CC 72 GGRVEEFLSKDISYLSNKKKAKFA-----OTLRISPPSPS----- 107
 CC 479 GPRIKMLNKLICDVLTKNPRFATRGPHPSQCNLFVVRDITFSGHSENFLEKMM 538
 CC 108 -----PESAVYAEETSPHPSH-----DSSSKSPDYTC----- 135
 CC 539 ALYVSHKKNPNQIMSNAPAHKFLVLPPIIDKDCG--RIPDPICVQIALGEISK 596
 CC 136 -----LSRG-----LLVERAKHDHDFPSMS-----ILSN-----ALMGVYILTI 172
 CC 597 ESVRNSLSRGRAGDILPWLISQFODEFASLSGAIYAIATNPYASMGSGAIEL 656
 CC 173 DDIRYITQKKELYLKSSSTVPRGC--KRVSGAQKTRTRQLKPKVYKEDMSQIX 229
 CC 657 --LRDYFGKFTDM-----SEVPRKDYSIKRV--SKELAKTMLAMD--VKLRDKTIL- 705
 CC 230 RPYTLQLTNMP-----FTNSIQKPCSPFVDKPSMOKOTQVLRITQDCKYG--GT 281
 CC 706 PLLKLKLEQPPHYLHYGVSYGLTSLHKF--WRNNSF--VPEYLR--QTANDLTGEHT 759
 CC 282 STIOLKKEKKKKGCCCLQKYEDLETHLSEQHNFNAQSNQYOVVDIVSKLVPDFVEY 341
 CC 760 CVMNLVLEGRESNMLVEFAKDFPKRFLSLSYDFHKFTAYVALSVIES--SKRAQDLSLD 817
 CC 342 EKDTPKK-----KRI-KYS-----VGSISPASAVLKKTQCKREVEL 377
 CC 818 EKHDKKELTRHLDIFSPDLKRLDSYGNLLDYHVGIMIPMLALLYFGDKMGDSVKL 877
 CC 378 QHISQK-----DCQEDDTVKEONFLYKKEOEKKULITSEPIHPNSNELRGINEM 430
 CC 878 SSVOGAIILAIQLOKRIIDITAKELH-----PNSQITAFPAKI 916
 CC 431 SKKSGMLSTAEEDIRQNFQPLAKKQCCIDISEHLSLENDLELRVDHYKNCIOASY 490
 CC 917 MKRMS-----QYFROL--LSQSTIEELTPKIDDDIAEDGEBELK-----NINAAE 959
 CC 491 HVSDESTD--NSGQAPKQSDVLFPAKDLKEDKLHSFTFHDSSGLITINS-----SQEHL 543
 CC 960 ALDQDEDELEFAGSEAVQ-----AMREKQKELINSLNDKKAINDNSEMAEASQKSL 1011
 CC 544 TVQAKAPFHTPEEPNECDPKNMDSLPSGK-----IHKKYKILIGRRKRENEPN 593
 CC 1012 EIAAKA-----KGVASLTKGKRTTEKADIYIOEKMAKKRPRKSKAAN 1056

RESULT 8

CENE_HUMAN STANDARD; PRT; 2663 AA.
 AC 002224;
 DT 01-JUN-1993 (Rel. 26, Created)
 DT 01-JUN-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DT CENTROMERIC PROTEIN E (CENP-E PROTEIN).
 DE CENE.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.,
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RT Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RA MEDLINE=95196755; PubMed=7889940;
 RA Thowar D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RT EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RA MEDLINE=96437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
 RT J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -1- CELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 215005; CAA78727.1; -
 CC DR PIR: S28261; S28261.
 CC DR HSSP: P03069; 1211.
 CC DR MM: 117143; -
 CC DR InterPro: IPR001752; Kinesin.
 CC DR Pfam: PF00225; Kinesin; 1.
 CC DR PRINTS: PRO0380; KINESINHEAVY.
 CC DR SMART: SM00129; KISC; 1.
 CC DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC DR PROSITE: PS50067; KINESIN MOTOR DOMAIN2; 1.
 CC KW Motor protein; Cell division; ATP-binding; Colled coil; Mitosis;
 CC Cell cycle; Centromere;
 CC KW DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).
 CC FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 CC FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 CC FT NP_BIND 86 93 ATP (BI SIMILARITY).
 CC FT SEQUENCE 2663 AA; 312087 MW; CEF1380C8C8C88 CRC64;

Query Match 4.4%; Score 154; DB 1; Length 2663;
 Best Local Similarity 18.2%; Pred. No. 1.5;
 Matches 147; Conservative 117; Mismatches 243; Indels 300; Gaps 37;


```

OY 18 IQVKNKRNPSLSKLTQNR-----PEKSKCPYMGKVVYLDLPSTVTSKQKQKIDKG 72
DB 1379 IQESQSKQDSINMEKKNETTKIYSEMEQFPKPSALIRIEMIGLSKRLQESHDEM- 1437
OY 73 GRVEEFLSKDISILSNKKEAKFAQTLGRISVPSPESAVTAETTSPHSHSGSSFKSPD 132
DB 1438 -----KSYAKEKDDLRLOEVLOSQSOLKENI----- 1465
OY 133 TVCLSRGKLVEKAIKDHDFPSNSILSNALSQVYKILH-----IDIRYYIEOKK 183
DB 1466 -----KEIYAKHLEETE-----ELKVAACJLQEQEETINELRVNLSERE 1504
OY 184 KELYLKSSSVSRD-----CGKRVGSAQCTFRGRLKPPYKVEDMS- 226
DB 1505 TEISTIQOLEKINIKLQNKIOEIEYKEEQLNKOISEVQEVNVELKQPKERKADSL 1564
OY 227 QLYRFYQLTNMPPIYNSIQKSPFDY-DKPSMOKOTQVRLRIOTDGYGCTSIQL 285
DB 1565 QSIESKMLELN-----RLQESQEIQIMKEKEMKRVQALQIERDQLEKNTKEIYA 1618
OY 286 QLKERRKRGY-----CECCLOREYEDLETHLSEQ-----HRN 317
DB 1619 KMKESQEKYQPLKMTAVNETQEKMCIEHLK-EQEFQKLMLENTEFENIRLTQLLHEN 1677
OY 318 FQASQNOY-QYVDIYKSLYFVEYEEKDPKKRKIKYSGSLSPSASVLTKEQEKYE 376
DB 1678 IEMMSVTKERDILRS--VEETLKVERDOLKN-----LREITRLQEKQELK 1724
OY 377 LOHI-----SQRQCOEDPTTYKEQNFPL-----YK 400
DB 1725 IYHMLKHEQETIDKLRIGVSEKTEISMNQKDLHSNALKAQDLQIOEELRIAMHLK 1784
OY 401 EFOETKEKLL-FISEPIHPSN--ELRGLNEMSKNCS-----MSTAEDDIRQ--- 446
DB 1785 EQOETIDKLRIGVSEKTEKLSNMQDLHSNALKQELQELKANEHQLITLKQDVNETQK 1844
OY 447 -----NFTQPLRKNKQEC-----IDLI 464
DB 1845 KYSEMEQLKQIKQDSLTLKLEINENLQO-ELHENEEMKSVKERNDLRVEETLKL 1903
OY 465 SEHTLSEN-----DLE--ELR-----VDHYKCNL-OASVHVSDFSTDSNG 501
DB 1904 EBDQLESIOETKARDLEIQOELKTRAMLSKHEKENVDLREKISKITQIDQICDLK 1963
OY 502 SOPKQKSDTVLPFANDLEKDLHSIFTHDSGLITINSQEHLT--VOAKAPFHTPEEPN 559
DB 1964 SK-----DELQKKIQELQKELQILRVKED--VMSHKIKINEMQLKKOF-----EPN 2009
OY 560 ---EEDFKMMDLPSGKTH--RKVKILG-----RNKENLEPNAEFDKRTFEI----- 603
DB 2010 YLCKCEMDNFQI--TKKLHSELEIRIVAKERDEDLRIKESIKM-----ERDOFIATLRE 2062
OY 604 -----TOENRNICSSPQSLID 620
DB 2063 MIARDQRNHQVPEKRLSLDSQGHLMF 2089

```

```

RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE-Brain;
RX PubMed-11118617;
RA Caricasole A., Sala C., Roncarati R., Formentti E., Terstappen G.C.;
RT "Cloning and characterization of the human phosphoinositide-specific
RL phospholipase C-beta 1 (PLCBeta1).";
RN Biochim. Biophys. Acta 1517:63-72(2000).
[2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE-Brain;
RX MEDLINE-20225428; PubMed-10760467;
RA Peruzzi D., Calabrese G., Faenza I., Manzoli L., Matteucci A.,
RA Gianfrancesco F., Billi A.M., Stuppia L., Palka G., Cocco L.;
RT "Identification and chromosomal localisation by fluorescence in situ
RT hybridisation of human gene of phosphoinositide-specific phospholipase
RL C beta 1.";
RN Biochim. Biophys. Acta 1484:175-182(2000).
[3]
RP SEQUENCE OF 261-1216 FROM N.A. (ISOFORM B).
RC TISSUE-Testis;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RN Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 519-1216 FROM N.A. (ISOFORM A).
RC TISSUE-Brain;
RX MEDLINE-98290545; PubMed-9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC C ENZYMES.
CC -1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE
CC + H(2)O = D-MYO-INOSITOL 1,4,5-TRIPHOSPHATE + DIACYLGLYCEROL.
CC -1- COFACTOR: REQUIRES CALCIUM.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE), AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-BETA 1 IS
CC MEDIATED BY TWO G-PROTEIN ALPHA SUBUNITS, ALPHA-O AND ALPHA-11.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS
CC OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DB EMBL: AJ278313; CAB98142.1; -
DB EMBL: AJ278314; CAB98143.1; -
DB EMBL: AY004175; AAF86613.1; -
DB EMBL: AL137267; CAB70666.1; -
DB EMBL: AB011153; BAA25507.1; -
DB HSSP: P10688; 10AS.
DR InterPro: IPR000008; C2.
DR InterPro: IPR001192; PL_PLC.
DR InterPro: IPR000909; PL_PLC_X.
DR InterPro: IPR001711; PL_PLC_Y.
DR Pfam: PF00168; C2; 1.
DR PRINTS: PR00390; PHPLIPASEC.
DR PRODOM: PD001202; PL_PLC_Y; 1.
DR SMART: SM00239; C2; 1.
DR SMART: SM00148; PLCXG; 1.
DR SMART: SM00149; PLCYC; 1.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS50007; PIPLC_X_DOMAIN; 1.

```

PROSITE: P550008: PIPIC_Y-DOMAIN: 1.
 KW Hydrophobic; Lipid degradation; Transducer; Phosphorylation; Calcium;
 KW Alternative splicing.

FT DOMAIN 316 467 DOMAIN X.
 FT DOMAIN 540 656 DOMAIN Y.
 FT DOMAIN 663 761 C2 DOMAIN.
 FT ACT_SITE 331 331 BY SIMILARITY.
 FT ACT_SITE 378 378 BY SIMILARITY.
 FT MOD_RES 887 887 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT VASAPLIC 1142 1216 LOVELEODYODKRLPLELLEFVOEAKKISDSNHGSA
 PLSLSDPGKVNKHRTPSSEELGDIPEKFEPTPL -> GEG
 SSSLETCHEBDSVSPNFTPPNPOLAKW (IN ISOFORM
 B).
 MAGAQPVHALQKPVCSDSLKKGTFKVKWDD ->
 MSLGQIATRTILRLISDALIKKRTDLKS (IN REF.
 2).
 FT CONFLICT 1 34
 FT CONFLICT 189 189 L -> M (IN REF. 2).
 FT CONFLICT 203 203 P -> L (IN REF. 2).
 FT CONFLICT 216 216 L -> F (IN REF. 2).
 FT CONFLICT 221 221 P -> L (IN REF. 2).
 FT CONFLICT 266 266 L -> P (IN REF. 2).
 FT CONFLICT 309 309 P -> T (IN REF. 2).
 FT CONFLICT 320 320 Q -> R (IN REF. 2).
 FT CONFLICT 352 352 V -> A (IN REF. 2).
 FT CONFLICT 366 366 K -> R (IN REF. 2).
 FT CONFLICT 393 393 E -> K (IN REF. 2).
 FT CONFLICT 983 983 P -> S (IN REF. 1; CAB98143).
 SO SEQUENCE 1216 AA: 138566 MW: 6F4263D1A50C6FD1 CRC64;

Query Match 4.3%; Score 152; DB 1; Length 1216;
 Best local similarity 20.9%; Pred. No. 0.68;
 Matches 117; Conservative 74; Mismatches 186; Indels 184; Gaps 22;

DB 18 IOVKNKNNPSLKTNDNPEKSKCKPLMGKVFYLDLPSTISEKLOKDKIDGKRVEE 77
 774 ICLNRNRP-----LTPAVFVTEVKDVPPTVADVE 808
 78 FLKSDISY-----ISNKEKFAQTIGRISVPSPESA-----YT 113
 809 ALSNPFRVYNLEQFAKOLALTEDEEVKKADPGE-TSEASEKFTTPANGVNH 867
 114 AETSPHPSHDSFKSPDVLCSRGKILVEKAIKNDHFIPSNITLSNALSNGVKLIH 173
 868 TPLTPPSPSALHSOPAGS-----VKAPAKTEDEL-QSVLE-----VEAQTE 911
 174 DIRVTEOKKKELEYLLKRSSTVSVDGKRVGSAQKRTGRILKRPVVEDMSQLYPER 233
 912 ELK-----QOKSFVKKQKKHKKEMKDLVR-----HNKRTDILKEHTTAYNEIONDY 959
 234 LQLTNMPFNVISIQ-----PCSPDVDPKSSMOKQTVKRLIQTDGKRYGCTSIQ 286
 960 --LARRAALKSAKDSKSKSEPPSP--DHGSSTEEDIALALDAEM-----TOKLID 1007
 287 LKEKKKGCCECLOKYEDELTHLLSDHNRFAOSNOYQVVDIVSKLVFDFVEYKOTP 346
 1008 LKDKOQ-----QQLNLNRQ-BOY-----YSEKIQ 1030
 347 KKKRIKSVGSLSPVSASY-----LKTEQKKEVLEOHISOKDQEDDTVEKQNFY 339
 1031 KREHIKILKILDLVADECONNLKRLKEIEKKEKELKKMKDKKQKQKQKQKQKQ 1090
 400 KETQETEKRLFTSEPIPHPSNELKGLNKKNSKSMSTAEDDIRQ-----446
 1091 MEERKTEMISYIOEVVQY-----IKRLEAOSKQKQKELVKEHNEIQOILDEKPKLOVEL 1146
 447 -----NFTOLPHAKKQCCILDISHTLSENDLE-----LRVDHYKCNQIASVH 491
 1147 BOEVODKFRRLPLE-----ILFVOEAMKGRKISDSNHGSAPLSLSDPGKVN-----H 1195
 492 VSDSTNSGSGPKOKSDTVL 512
 1196 KTPSEELGDIPEKFEPTPL 1216

RESULT 10
 RAD50_YEAST STANDARD: PRT: 1312 AA.
 AC RAD50_YEAST
 AD P12753;
 DC 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 CN DNA REPAIR PROTEIN RAD50 (153 KDA PROTEIN).
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4912;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC SPRAIN-RE821;
 CC MEDLINE=89276917; PubMed=2659437;
 CC Alani E., Subbiah S., Kleckner N.;
 CC "The yeast Rad50 gene encodes a predicted 153-kD protein containing a
 CC putative nucleotide-binding domain and two large heptad-repeat
 CC regions.";
 CC Genetics 122:47-57(1989).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC SPRAIN-S288C / FY1679;
 CC MEDLINE=97377992; PubMed=9234673;
 CC Sen-Gupta M., Guedener U., Behnhauer J.D., Fiedler T.A.,
 CC Behnhauer J.H.;
 CC "Sequence analysis of the 33 kb long region between ORC5 and SUI1
 CC genes, the left arm of chromosome XIV from Saccharomyces cerevisiae.";
 CC Yeast 13:849-860(1997).
 CC [1]
 CC FUNCTION: IS REQUIRED DURING MEIOSIS IN THE PROPHASE FOR
 CC CHROMOSOME SYNAPSIS AND HOMOLOGOUS RECOMBINATION. DURING
 CC VEGETATIVE GROWTH IT IS NECESSARY FOR DNA REPAIR. THE PROTEIN MAY
 CC BE INVOLVED IN SEARCH FOR HOMOLOGY BETWEEN INTERACTING DNA
 CC MOLECULES OR CHROMOSOMES. COULD BE A 5'-3' EXONUCLEASE.
 CC WORK IN COMPLEX WITH MEI1.
 CC [1]
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC [1]
 CC EMBL: X14814; CAA32919.1;
 CC EMBL: X96722; CAA65494.1;
 CC EMBL: 271526; CAA96157.1;
 CC PIR: S05808; BMBYDL.
 CC SGD: S0005194; RAD50.
 CC InterPro: IPR003439; ABC_transport.
 CC InterPro: IPR001687; ATP-binding; Metasts.
 CC Repeat: DNA repair; ATP-binding; Colled coil; Metasts.
 CC NP_BIND 34 41
 CC DOMAIN 185 347 COILED COIL (POTENTIAL).
 CC DOMAIN 403 558 COILED COIL (POTENTIAL).
 CC DOMAIN 617 672 COILED COIL (POTENTIAL).
 CC DOMAIN 734 1108 COILED COIL (POTENTIAL).
 CC SEQUENCE 1312 AA: 152568 MW: 58A0AAL73AC5677E CRC64;

Query Match 4.3%; Score 152; DB 1; Length 1312;
 Best local similarity 18.0%; Pred. No. 0.75;
 Matches 144; Conservative 151; Mismatches 288; Indels 215; Gaps 34;

DB 8 IHKGFPGGIGOVKNKNNPSLKTNDNPEKSKCKPLMGKVFYLDLPSTISEKLOKD 67
 345 IRRQGLEAGKET-YEKRNHLISLK-----EAFQHK-----FQGLSNITNSDAQVNA 391
 68 -----IKDLGRVEEFLSKDISYLSNKKKFAQTIGRISVPSPESAATYAEFT 117

Db 392 HEMGOFKAFISODLTDIDOF-AKDIDL-----KFTNLSDLIKSTT-VDSONLEYNKADR 444
Qy 118 SP--HPSHDGS-----SPKSPPTVCLSRGKLLVEAKIXDHPFSPNSILSNALSMGVKILH 171
Db 445 SKLHDSLELAKELKSPFSLST-----QDSLNHELENLKTVEKYLQMSSEMI- 492
Qy 172 IDDIRYIEQKKKELYLLKSTSTSVROGKRVGSGAQ-----KRTGKLKPF 219
Db 493 IPKLNQKLEENKNNEMIILENOIEKFORIMKTNQADLYAKGLIKKSINIKRLDELQKIT 552
Qy 220 VKVEDMSQLYPEFY-----OLTMNPFIN 243
Db 553 EKLQDSIRIOVFPLTQEFQADLEMDFOKLFIMQKNIAINNNKMHLEDRRYTALINL 612
Qy 244 YSIOKPCS-----PFD--VDKPSMOKOTQVKLRIDTGDGKYGTSIQ 284
Db 613 NTIEKIDQDNQKSEKVIQQLSENPEDCITDEVNDVLEETELSKYKALENLKMHTLE 672
Qy 285 LQ---LAKKKKGCCECLQKVEDLETHLLSEQHNFQASNOYQVVDIVSKLVDPVEY 341
Db 673 ENRRALIEAERDSCCYLCSRKEFN-----ESFESKLLQELKTKTDANFERT 718
Qy 342 EKDPKKR-----IKYSGSLSPVASY--LKTEQKKEVLOHISOK--DCOEDD 389
Db 719 LKDIYQNKETLHSLRLLEKHTITLNSINEKIDNSOKLEKAKETKTSKSLDELEVD 778
Qy 390 TTVKQNFLEYKETOETEKLLFISEPIPHSPNELRGL--NEKMSNKCMLSTAD-- 442
Db 779 TKLKE-----KELASEIRPLI--EKFTYLEKELKLENSKTSIEELIYTSDEGIQT 832
Qy 443 -----DIRONFOTPLKKNQ-----ECILDISHTISENLEELRDH 481
Db 833 VDELNDQQRKNDLSIRELAKTISDQMEKDEKVRNSRMINIKELKELVSEIBSLQ- 891
Qy 482 YKCNQASVHVSDFSTDSGSGOPKSDTVLEPAKDLKEDLHSTFTHDSGLITINSOE 541
Db 892 -KONIDDSIRSKREINIDISRVKELARIT--SLKKN-----KDEASVLDKAKN 939
Qy 542 HLTVQAKAP-----FHTPPEPNEDCFKNDSPSGRIHRVKITILGRNKE 588
Db 940 ERDIOVRKKQKTVADINLIDRFQITVNEVDFEAKGFDELQTT--TIR-ELELNKQ 993
Qy 589 NLEPNAEFDKRTETFTQEBNRICSS--PVQSLDLFQTSSEKSEFLGFTSYEKGICNV 646
Db 994 MLELKEQDLKSNVEYNEKRKLADSNNEKNKQNLIELKLSQLIHSEISRIQVQNA 1053
Qy 647 ---LDIMEENDNLTA 662
Db 1054 EAERDKQOE-SLRLLTIR 1071

RESULT 11
MY9B_RAT MY9B_RAT STANDARD: PRT: 1980 AA.
AC Q63358;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN IXB (UNCONVENTIONAL MYOSIN-9B).
GN MYO9B OR MYR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
RX MEDLINE=95188874; PubMed=7882973;
RA Reinhard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
RA Baehtler M.;
RT "A novel type of myosin implicated in signalling by rho family
RT GTPases.";

RL EMBL J. 14:697-704(1995).
CC -I- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC -I- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
CC PERINUCLEAR REGION (BY SIMILARITY).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
CC LIVER, AND SPLEEN.
CC -I- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHOSPHO-ESTER AND DAG
CC BINDING DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 GAP DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77609; GAA54700.1; -.
DR HSSP: P08799; 1MND.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR00159; RA.
DR InterPro: IPR00198; RhogAP.
DR Pfam: PF00130; DAG_PE-bind; 1.
DR Pfam: PF00612; IQ; 4.
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00620; RhogAP; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 2.
DR SMART: SM00109; C1; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MSC; 1.
DR SMART: SM00314; RA; 1.
DR SMART: SM00324; RhogAP; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin. Repeat. ATP-binding. Calmodulin-binding. Actin-binding;
KW Cytoskeleton; Coiled coil; GTPase activation; Phospho-ester binding;
KW Zinc.
KW DOMAIN 1 940 HEAD OR MOTOR DOMAIN.
FT DOMAIN 941 1045 NECK OR REGULATOR DOMAIN.
FT DOMAIN 1046 1980 TRAIL.
FT DOMAIN 1841 1861 COILED COIL (POTENTIAL).
FT DOMAIN 1918 1948 COILED COIL (POTENTIAL).
FT DOMAIN 845 856 ACTIN-BINDING.
FT DOMAIN 958 978 IQ 1.
FT DOMAIN 981 1001 IQ 2.
FT DOMAIN 1002 1024 IQ 3.
FT DOMAIN 1025 1054 IQ 4.
FT DOMAIN 1593 1641 PHOSPHO-ESTER AND DAG BINDING.
FT DOMAIN 1673 1822 GAP DOMAIN.
FT NP_BIND 239 246 ATP (POTENTIAL).
FT SEQUENCE 1980 AA: 225035 MW: D79FECA0FADFAC05 CRC64;

Query Match 4.3%; Score 152; DB 1; Length 1980;
Best Local Similarity 20.0%; Pred. No. 1.3;
Matches 136; Conservative 107; Mismatches 242; Indels 196; Gaps 33;
Qy 6 MRHS--KGHPQGIQVKNKKNRPSLKLTDNRPKSKCKPLMGVYFDLPSTVI--- 60

```

Db 1030 IRLDLSLGHQ-----RRSFQOMLEKOKAQOANETGAGMSGCE-----PSVAAE 1078
OY 61 --SEKRLDNDLQGRVEEFLSKDISYLSNKEEAKFAOTLGRISPPSPESAYTAETS 118
Db 1079 QSEHPVADPESLQVETETWMMNSKSPNGISPKKE-----IPSEMTTPAKQTV 1126
OY 119 PPHSDGSSFKSPQVQSLSGKLLVEKAIKHDFIPSNLSLNSMGVYLIHIDIRY 178
Db 1127 PAESHE-----KVPS-----SREK-----RRSRQRCLE-----H 1151
OY 179 IEOKKELLYLKSSSTVVRGCGKRVSSGAKTRQRLKPKPVK-VEDMSQLYRPFYLOLT 237
Db 1152 VEROKKHIIQSGRENSLGRPRSKASLETGESFPEDTKEPREDELFTWTETAPSCPK-- 1209
OY 238 NMPTLINSIQKPCSPFYVYKSPSMQKQYV-----KLRIQTGDKYGGTSTQLOLAKK 290
Db 1210 OVPYVG--DPPSPSPLOQFASLDLSRVSPVLPSSLSLSPQDEDK-GENSTKVVDKPE 1265
OY 291 KKKGYCECCLOKYEDELTHLSE-----OHRNPAQSNQYVVDIYSKLVDFVEYKND- 344
Db 1266 SPFSSTQ--IQRYOHPDTERLATVAEIMWGKKLASAMLSQSLD-----LSEKPR 1312
OY 345 -----TP-KKKRIKYS--VGSLSFVSASVLYKTEQKEKVELQHIQND----- 384
Db 1313 TAGAALPTEERRRISFSTSDVSKISPVKSTFVDDGLSAPKPKAGKKKSSDPGAGDAGL 1372
OY 385 ---COEDDTYKQNFLYKFTQTEKKLFISEPIPHPSNELRGINKMSNKCMLSTAE 441
Db 1373 PTGSQGDERSAFKRLFLHK---AKDK-----PS-LEGVE-----TE 1406
OY 442 DDIRONFQULPHKNOCELIPI-----SEHTLSENDEELRVHYK---CNLQASVHS 493
Db 1407 GSGGQAQGEAPAKT-----LDVPSQGHRTTGKPLKGGKNNRNVGQITVSEKRRS 1461
OY 494 DPSTDNSSQPKOKSDTVLPAPADLKEKDLHSIFTHDSGLITINSQEHLYQAAPRFT 553
Db 1462 VFRKTTANLELKLDFEFLNKNVNDLSQ-----KTPLESFILATPRFAS 1506
OY 554 PPEENEDPFRNMDSLPSCGIHKVYIIIGRRKNLE---PAAEFPKREFFTQENNI 610
Db 1507 -----NITMTSVNGKIHVGTIDLM-----ENVQIVVSNLAER-----GEMQML 1548
OY 611 CSSPVQSLDLFQTSSEKSEF 631
Db 1549 VLVNFQSLDDEFRTSYNKTFD 1569

RESULT 12
YAF3_SCHPO STANDARD: PRT: 1044 AA.
AC Q09857: Q9P7V0;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DN HYPOTHEICAL I19.1 KDA PROTEIN C29E6.03C IN CHROMOSOME I.
GN SPAC29E6.03C OR SPAC30.07C.
OS Schizosaccharomyces pombe (Pisition yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCB1_TaxID=4896;
OY NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Jones L., Murphy L., McNeill A., Simpson I., Harris D., Barrell B.G.,
RA Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Brown S., Murphy L.,
RA Jones L., McNeill A., Harris D.;

```

```

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -I SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: Z66525; CA91425.1; ALT_SEQ.
DR EMBL: AL136338; CAB66466.1; -
KM Hypothetical protein; Coiled coil.
FT DOMAIN 600 1014 COILED COIL (POTENTIAL).
SQ SEQUENCE 1044 AA; 119127 MW; 52ACEDE174725A59 CRC64;

```

Query Match 4.38; Score 151.5; DB 1; Length 1044;
 Best Local Similarity 19.58; Pred. No. 0.6; Indels 201; Gaps 34;
 Matches 130; Conservative 122; Mismatches 214;

```

OY 52 YLDPVYISSEKLOKIDKLGGRVEEFLSKDISYLSNKEEAKFAOTLGRISPPSPESA 111
Db 418 YLDE--VHLSDDQDYD-----NMFTSTLYTVLIDNQRRYL-----LCSIFLQDM 462
OY 112 YTAETTPHPSHDGSSFKSPDYVCLSRGLVY-----EKAIKD----- 149
Db 463 DNDEDESE--SEDKVTF-----IQCVST-KLIATLRHENALQNCVGLVTLIALVYGNPDS 514
OY 150 -HDFIPNSIL-----SNALSMGVKLIHIDIRY--IEO--KKKELYLKK 191
Db 515 VKDLSSESLIQFLFALMDESSANSYIOGMIAVFLSLVYYICPIESPVSQDY---- 570
OY 192 SSVVRGCGKRVSSGAKTRQRLKPKPVYKEDMSQLYRPFYLOLNMFFIYVSIQKPCS 251
Db 571 -----NAITSVAKRD-VFINRLQRLR--NMLVEIFFL----- 600
OY 252 PFVDKSPSMQKQYK-RIQTDGDKYGGTSTQLOKE-----KKKGYCEC 298
Db 601 -----SMQKQKLSLREIDNTR--EALDSYKENSIOBEKLNLSLSTKTNDLE 649
OY 299 CLQRYEDELTHLSEQHR-----NPAQSN-----QYQVVDIYSKLVDFV-ETEK 343
Db 650 QTOIAEKYHELLDNOQKLYDIRIELDYTKSNCKOMEQVLRGEHSEIADFIEHSK 709
OY 344 DTPKKRIKYSVGLSPVSASVLYKTEQKEV--ELQHIQKDCQ--EDDTYKQNFLYK 400
Db 710 LTKQLDIDIKNOFGIISKNRDLISLEKSKSLNNSLALESKNKKLENDLILTER-LNK 768
OY 401 ETOETEKLLFISEPIPHPSNELRGINKMSNKCMLSTAEEDIRONFT-----QLPL 453
Db 769 KNAETES-----FKNTIREAELESKKALNDNLGNKEITISDLKNKLSSESTRLOEQSLQND 824
OY 454 HKNOCELIPI-----ISEHTLSENDEELRVHYDK-NIQASVHSDSTNSQSO 503
Db 825 DKNOLETNLRIAAADELSSMESLNKNQANLEKLAOKCSNLOEKI-----NFGNK 876
OY 504 -PKQSDTVLPFAKQDK-----EKDISFTHDSGLITINSQEHLYQAAPRFT 553
Db 877 LAKHTYKISLKLDEAATKTAASLSKELTVKSENDLSKVSNDQK----- 925
OY 554 PPEEPNECDTRKND-----SLPSGKIHR-KVITIIIGRRKNLEPNAEFKREF 602
Db 926 KEKSVNNEKREVSQALAEANLEKLNARDEITERLKVDTIGLQNASINQSLDSDSNRTIS 985
OY 603 ITOEENR 609
Db 986 DLESKNK 992

```

RESULT 13
 MY3_SCHPO

KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is not in
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X03831; CAAT7446.1; .
 CC PIR: A23386; SAZOGM.
 CC InterPro: IPR000561; EGF-like.
 CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC KW SIGNAL 1 19 POTENTIAL SURFACE PROTEIN 1
 CC FT CHAIN 20 1726 MEROZITE SURFACE PROTEIN 1
 CC FT CAROBYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1121 1221 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRC64;
 SO

Query Match 4.3%; Score 149.5; DB 1; Length 1726;
 Best Local 20.4%; Pred. No. 1.4; Mismatches 224; Indels 223; Gaps 37;
 Matches 141; Conservative 104; Mismatches 224; Indels 223; Gaps 37;
 65 QDIDSLGGRVE-----FLSKDISYLSNKKKFAKNOTGRTSPVPSPESAVYAFETS 118
 309 QYDLSTYKKOLEAHNLISVLEKRIOTL--KKENIKELLDKIN-----EIKN 354
 119 PPHSDGSFSPPTVLSRGKLLV--EKAKIDHDFIPNSILSNLSMGVKILHID-D 174
 355 PPHASGN--TPNTL-LDKNKKEIEHEKKE-----IAKTKFNIDSLFDPLE 401
 175 IRYIEOKKKELEYLLKSSVSVDGKRVGSAQKTRGLKKEFEVVEDMSQLRPYL 234
 402 LEYLKRNKKKVDYPKSODPKS-----V 426
 235 QLTNMPFINYSIKRCPSPFY-----DKPS-----SMOKOTOYKL-RIOTDGDK--- 277
 427 QILKVPYFN-GIVYPLPLDILNSLADNDKNSYGDIMNPOTKEINKEIITNKKRIE 485
 278 YGSTSIOQLKKEK---KKGCECCLOKYE---DLETHLSSEGRNPAOSNGYAVDDI 330
 486 INNKKIDOLEKKNHNRKEDNKKLEDEYKSKDYBELLEFYEMFNNPNPKVDYDKI 545
 331 VSKLYVDFEYIEKDPKKK-RIKYSVGSISPYASATKTBKKEVLEHISQDCQEDD 389
 546 FS-----ARTYNEKORYNNKSSNSNYSYVQKLKA-----LSYLEDYSLRKG- 591
 390 TYVKEONFLYKETOET--EKKLEFISPIPHSPNPLRGLENNKSKCMILSTAMDDIRON 447
 592 --LSEDFEHHYTLTGLLEADIKITBELIKSSNK-----LEKN 629
 448 FTQPLAKKOCCLIDISEHTLSF-----NDLEPLRDYHKKNIOASVYSPFSDNS 500
 630 FKGLTHSANASLEVIDYKLVQVLLKRIEKLRIELFLKNALAKNSIHVPITYPQK 669
 501 GSPD-----KKSPTV---LPPAKDLKEDD---LSTI-----FTHSGILITNS 539

DB 690 KPEPYYLVKKEVDKLEFIPKVKMDLKEQAVLSITOPLVASSETEDGSHYTLSS 749
 QY 540 Q-----EHLTVQAKAPFT-----PPEE-----PNECDPKMDSLPGKIKRRVKII 581
 DB 750 QSEFEVTEPEPEPEPEVGHVYITLPLPKKVVENSIRKNSDN----- 796
 QY 582 LGRNRENEPNAEEDKTEPITQENRIC-----SPVOSLIDLFQ-TSEKSEEL 632
 DB 797 -----SQALTIVYKRIKDEFLT--KSYICHKXIIYVSSMDOKLLEVYLTPEENEL- 848
 QY 633 GFTSYEKSGICNVLDIWEENSNDLTAFFS 664
 DB 849 -----KS--CDPLDL--FNIONNIPAMYS 869

RESULT 15
 MSPL_PLAEP STANDARD; PRT: 1726 AA.
 ID MSPL_PLAEP
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZITE SURFACE PROTEIN 1 PRECURSOR (MEROZITE SURFACE ANTIGENS)
 GN (PMMSA) (GPI95).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate Palo Alto / Uganda).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_Taxid=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=8900525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the
 RT major merozoite surface antigen (gpi95) of the Uganda-Palo Alto
 RT isolate."
 RL Exp. Parasitol. 67:1-11(1988).
 RL -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MEROZITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is not in
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M37213; AAA29611.1; .
 CC InterPro: IPR000561; EGF-like.
 CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC KW SIGNAL 1 19 POTENTIAL SURFACE PROTEIN 1
 CC FT CHAIN 20 1726 MEROZITE SURFACE PROTEIN 1
 CC FT CAROBYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1121 1221 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT CAROBYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL)
 CC FT SEQUENCE 1726 AA; 196174 MW; 5B59CEBEFA2F9A026 CRC64;
 SO

THIS PAGE BLANK (USPTO)